The New NOAA R & D High Performance Computing System (RDHPCS)

Jet Management Team

1 August 2006

Brief History

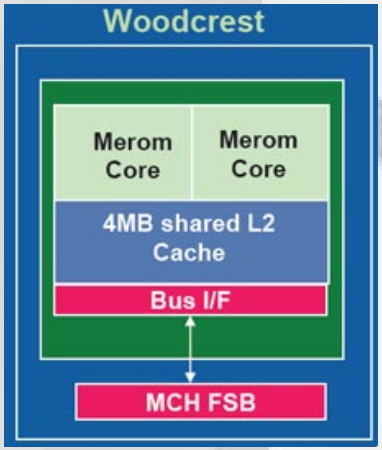
- Three Years Ago: Began Upgrade Procurement
- Two Years Ago: DOC and NOAA Mandated a (new) consolidated R & D HPCC procurement
  - ESRL/GSD
  - GFDL
  - NCEP (exclusive of operational component)
- May 2006: Announced Raytheon as the new integrator
  - SGI Altix in Princeton
  - IBM in Gaithersburg
  - Appro Linux Cluster in Boulder

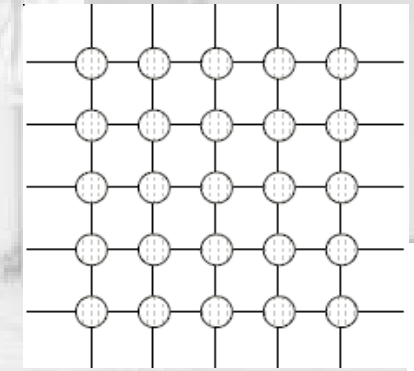
Status

- eJet will be on partial maintenance
  - We expect a large portion of eJet to be available throughout the year
- iJet will be off maintenance
  - We expect iJet to be mostly functional for the year
- HSMS (MSS) will be under maintenance for the year
  - With ample tapes
- New system installs mid-September
  - Acceptance: 10/2
  - Operations: early November*

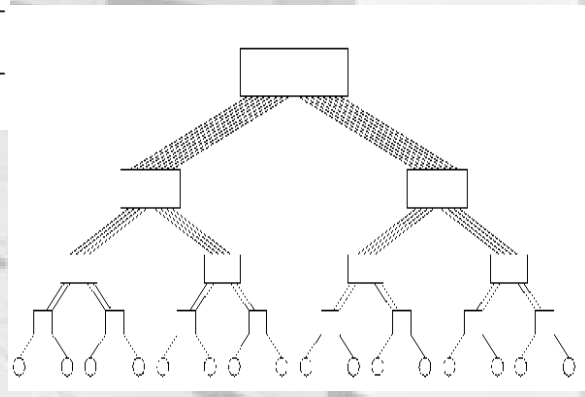
Basic Features of the New Boulder System

- Intel "Woodcrest" chip
  - Next generation Xeon
  - Dual-core/Dual-Socket
  - 2.667GHz
- Infiniband Interconnect
  - Mesh topology
  - (iJet and eJet are Myrinet "fat trees")
- Linux O/S
- SGE Batch System
- Terrascale File System
- ADIC HSMS
- New Machine Room (GA405)





Mesh

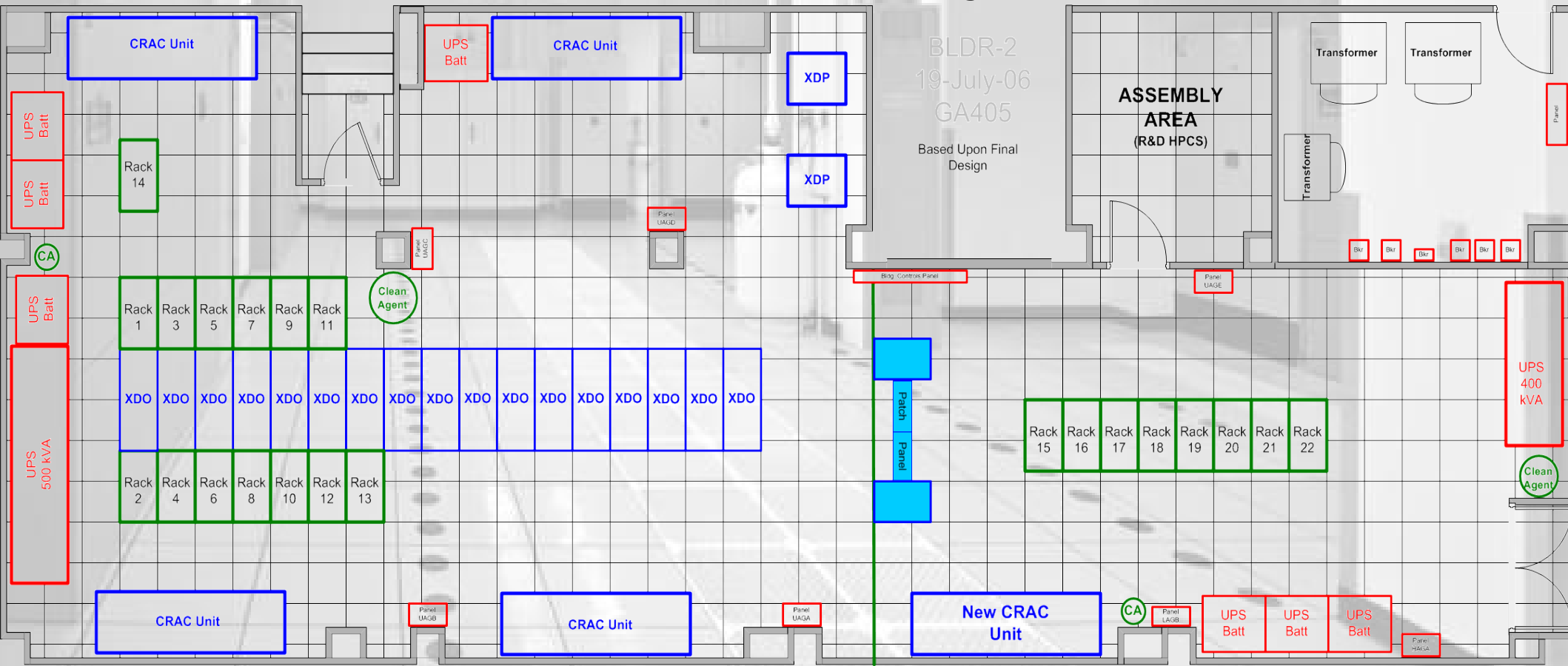


Fat Tree

noaa research
NOAA'S OFFICE OF OCEANIC AND ATMOSPHERIC RESEARCH

GA405

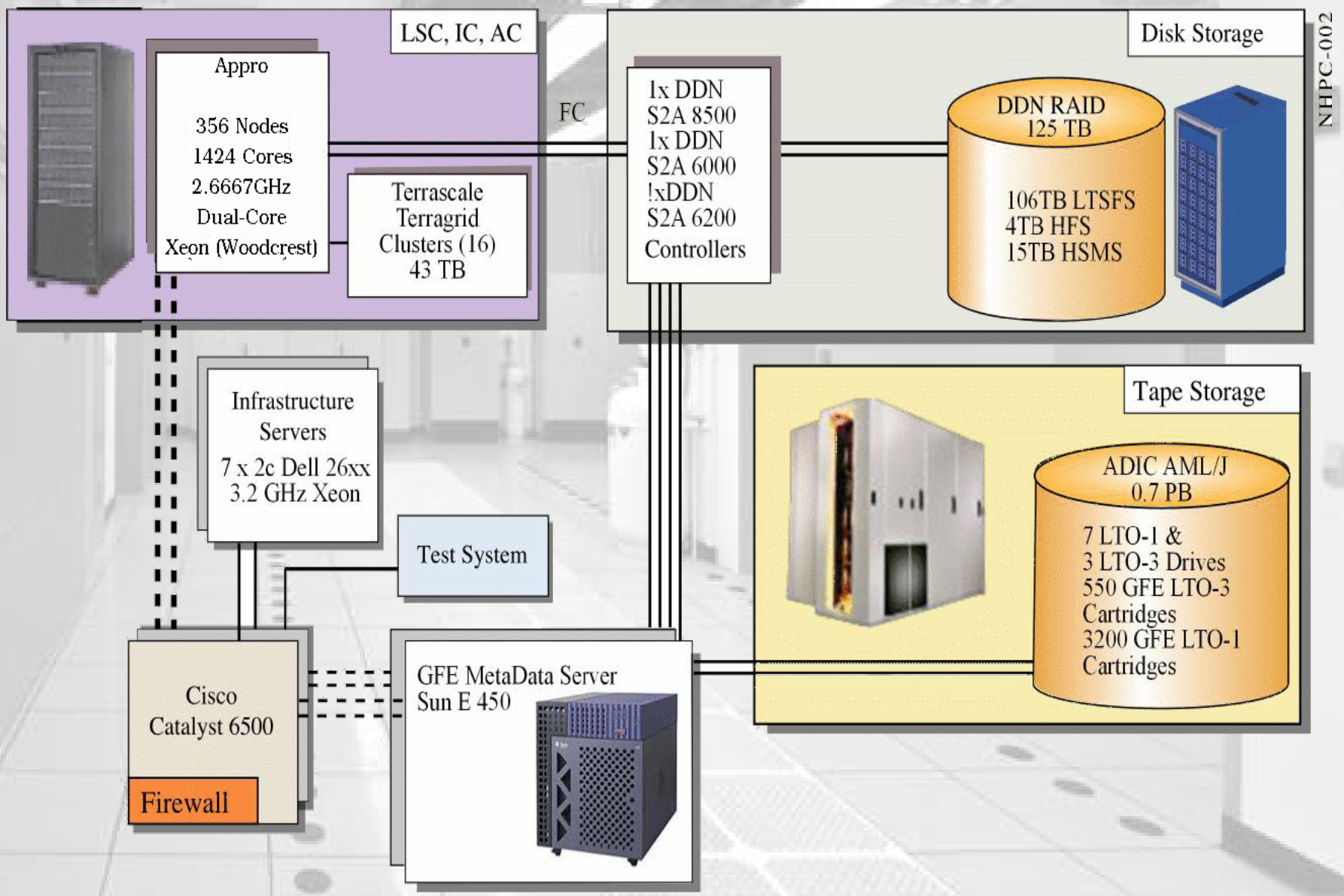
- Built for high density equipment
- 2200 square feet
- 400 KVA
  - 350 KVA East Side
  - 50 KVA West Side
  - NO Generator backup
- Monitoring
  - Orderly shutdown if necessary
  - Automated monitoring of CRACs, UPSs, etc
  - Visual monitoring with four web cams



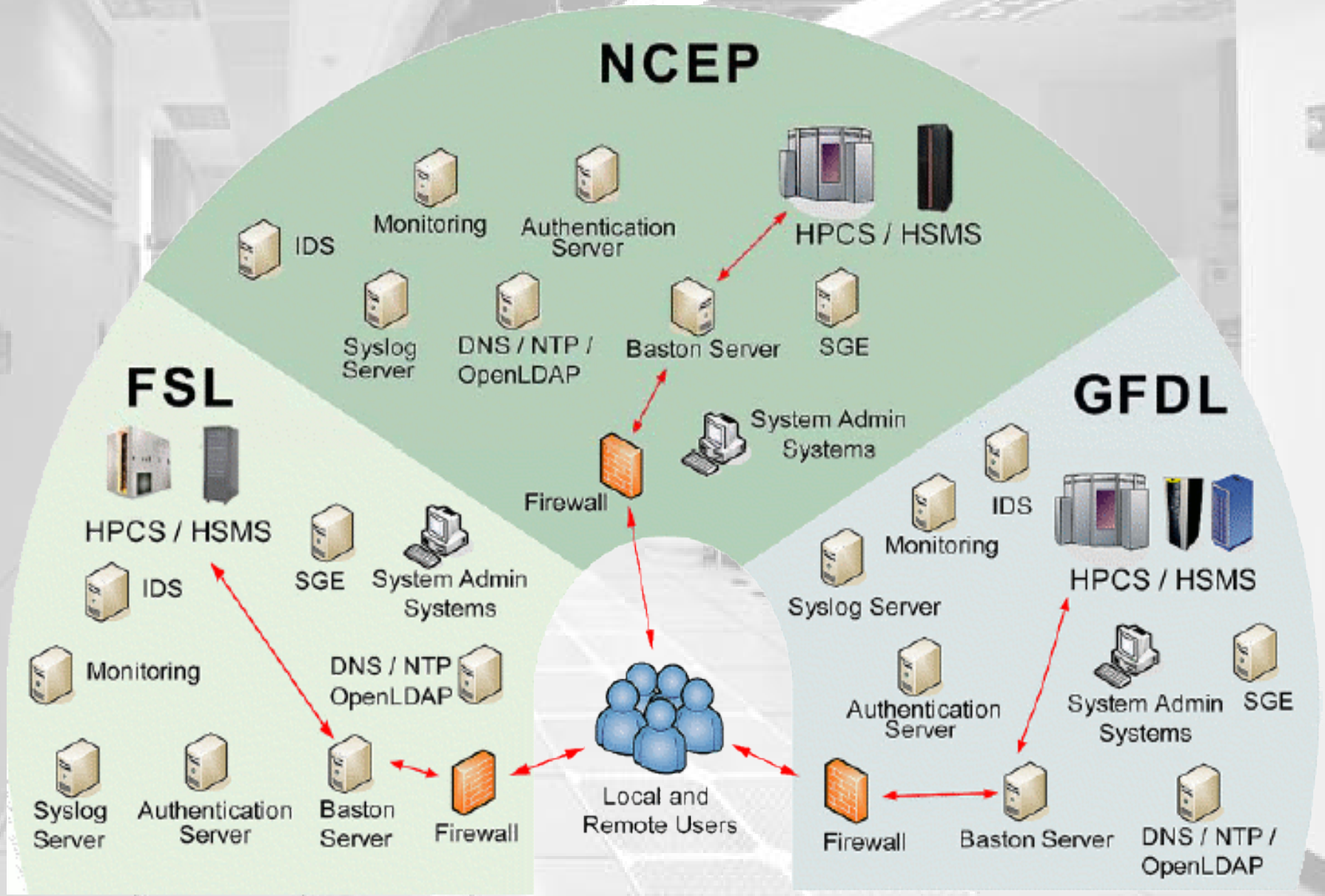
noaa research
NOAA'S OFFICE OF OCEANIC AND ATMOSPHERIC RESEARCH

GA405

Basic Features



Merging Toward "One-NOAA"



Cluster

- Cores
  - 356 Nodes
  - 1424 Cores
- Memory
  - 340 Nodes with 4 GBytes/Node (1 GByte/Core)
  - 8 Nodes with 8 GBytes/Node
  - 8 Nodes with 16 GBytes/Node
- Miscellaneous
  - Two front-end nodes
  - Intel and PGI Compilers
  - TotalView Debugger (limited amount of licenses)
  - 20 I/O Servers (Terrascale)

Disk Storage

- Three File Systems – All served with Terrascale
  - Home (HFS) – around ~4TB
  - Long Term Scratch (LTSFS) – ~106TB
  - Fast Scratch (FSFS) - ~43TB
- HFS will be used as it is today
  - Backed up, but limited storage
- LTSFS will be similar in function to the /pxx file systems
  - Large Storage, but not backed up (HFS+LTSFS bandwidth <2GB/s aggregate)
- FSFS will be a new feature for Boulder users
  - Purged very quickly, but faster (300MB/s per node and >3GB/s aggregate)
  - Used for intermediate files
  - Copy in/out from LTSFS

noaa research
NOAA'S OFFICE OF OCEANIC AND ATMOSPHERIC RESEARCH

Mass Storage System (HSMS)

- Not much has changed
  - ADIC StoreNext
  - ADIC AML/J robot
  - LTO-1 drives (7)
  - LTO-3 drives (3 – installed next week)
  - Access will still be through the mssXxx commands

Software

- SGE will not change significantly (if at all)
  - We may eventually integrate all of the systems under a single SGE instance
  - The basic resource name will be wcomp
    - qsub -pe wcomp 32
- The compilers and options will not change
  - There may be some new optimization options associated with the new chip set
- We will begin using “modules”
  - Modules will allow us to keep various revisions of system software (compilers, etc.) available
  - Instead of ***source /usr/local/bin/pgsettings-xxx.csh*** use ***module load PGI-xxx*** (detailed documentation will be provided as we implement)

Accessing the Systems

(Tentative)

- The domain name will change for all systems
  - rdhpcs.noaa.gov
- Access for all users (including GSD) will require the use of tokens
- Cluster Access
  - ijet.rdhpcs.noaa.gov
  - ejet.rdhpcs.noaa.gov
  - wjet.rdhpcs.noaa.gov
  - jet-scp.rdhpcs.noaa.gov
- Web Information
  - rdhpcs.noaa.gov/boulder

Resource Allocations

- For FY2007, resource management will be "status quo"
  - Allocations will continue to be handled through the Jet Allocation Committee
- Starting in FY2008, allocations will be handled through the PPBES process and administered by the Environmental Modeling Program (EMP)
  - Projects will encompass many sub-projects
    - Sub-projects are what we call "projects" on *Jet
  - Sub-projects will be administered by project managers
- Currently, there is a local Boulder discretionary allocation that will probably be managed by the Jet Allocation Committee

Transition

- Installation begins on 9/13/2006
- There will be outages
  - We will keep these as brief as possible
  - Both full and partial outages
- System B (the real-time system) will move to GA405
- System A (serial nodes) will move to GA405
- The DDN S2A8500 will be upgraded from 50TB to 120TB
  - /p70,/p71,/p72,/p73,/el0,/el1 will be migrated to the new LTSFS – this will take time (less if you remove unneeded files)
- Acceptance of wJet will begin on 10/2 with the system being generally available in early November (if all goes well)


noaa research
NOAA'S OFFICE OF OCEANIC AND ATMOSPHERIC RESEARCH
Thanks!
Questions?